ABSTRACT

The invention provides a new, efficient method for the assembly of protein tertiary structure from known, loosely encoded secondary structure constraints and sparse information about exact side chain contacts. The method is based on a new method for the reduced modeling of protein structure and dynamics, where the protein is described by representing side chain centers of mass rather than alpha-carbons. The model has implicit, built-in multi-body correlations that simulate short- and long-range packing preferences, hydrogen bonding cooperativity, and a mean force potential describing hydrophobic interactions. Due to the simplicity of the protein representation and definition of the model force field, the Monte Carlo algorithm is at least an order of magnitude faster than previously published Monte Carlo algorithms for three-dimensional structure assembly. In contrast to existing algorithms, the new method requires a smaller number of tertiary constraints for successful fold assembly; on average, one for every seven residues as compared to one for every four residues. The reliability and robustness of the invention make it useful for routine application in model building protocols based on various (and even very sparse) experimentally-derived structural constraints.